UNITED STATES UTILITY PATENT APPLICATION

FOR

THREE-DIMENSIONAL MEMORY ARRAY AND METHOD OF FABRICATION

Inventor:

N. Johan Knall

Attorney Docket: 003558.P007

Prepared by:

BLAKELY, SOKOLOFF, TAYLOR & ZAFMAN, LLP
12400 Wilshire Boulevard, 7th Floor
Los Angeles, California 90025-1026

EXPRESS MAIL CERTIFICATE OF MAILING

"Express Mail" mailing label number EL143570458US Date of Deposit 4/28,2000
I hereby certify that I am causing this paper or fee to be deposited with the United States Postal Service "Express Mail Post Office to Addressee" service on the date indicated above and that this paper or fee has been addressed to:

Box Patent Application
Assistant Commissioner for Patents
Washington, D. C. 20231

By: [signature] Date Signed: 4/28/00

THREE-DIMENSIONAL MEMORY ARRAY AND METHOD OF FABRICATION

BACKGROUND OF THE INVENTION

1. Field of the Invention.

The invention relates to the field of vertically stacked field programmable non-volatile memory and method of fabrication.

2. Prior art.

Recently there has been an interest in fabricating memories having memory cells disposed at numerous levels above a substrate. Each level includes a plurality of spaced-apart first lines extending in one direction which are vertically separated from a plurality of parallel spaced-apart second lines extending perpendicular to the first line. Cells are disposed between the first lines and second lines at the intersections of these lines. These memories are described in U.S. patents 5,835,396 and 6,034,882.

As will be seen, the present invention departs from the structures shown in these patents and uses "rail-stacks" as will be described later. The invented memory employs antifuses where a diode is formed upon programming a particular bit. In this connection see, *"A Novel High-Density Low-Cost Diode Programmable Read Only Memory,"* by de Graaf, Woerlee,

Hart, Lifka, de Vreede, Janssen, Sluijs and Paulzen, IEDM-96, beginning at page 189 and U.S. patents 4,876,220; 4,881,114 and 4,543,594.

SUMMARY OF THE INVENTION

A multi-level memory array disposed above a substrate is disclosed. A first plurality of spaced-apart rail-stacks disposed at a first height and a first direction are fabricated above the substrate. Each rail-stack includes a first conductor and a first semiconductor layer extending substantially the entire length of the first conductor. A second plurality of spaced-apart conductors are disposed above the first conductors and run in a second direction different than the first direction. An insulating layer is formed between the first rail-stack and the second conductors which is capable of being selectively breached by passing a current between one of the first and one of the second conductors to program the array.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a perspective view of a cut-away portion of the invented array.

Figures 2A-2H illustrate some of the steps used to fabricate one embodiment of the invented memory.

Figure 2A is a cross-sectional elevation view of an antifuse and semiconductor layer formed during the fabrication of the invented array.

Figure 2B illustrates the structure of Figure 2A after an additional semiconductor layer has been formed.

Figure 2C illustrates the structure of Figure 2B after a conductive layer is formed.

Figure 2D illustrates the structure of Figure 2C after an additional semiconductor layer has been formed.

Figure 2E illustrates the structure of Figure 2D after a masking and etching step.

Figure 2F illustrates the structure of Figure 2E after open spaces left from the etching step have been filled.

Figure 2G illustrates the structure of Figure 2F after a planarization step.

Figure 2H illustrates the structure of Figure 2G after another antifuse layer is formed.

Figure 3 is a cross-sectional elevation view of one embodiment of the present invented array.

Figure 4 is a cross-sectional elevation view of a second embodiment of the invented array.

Figure 5 is a cross-sectional elevation view of a third embodiment of the invented array.

DETAILED DESCRIPTION OF THE PRESENT INVENTION

A three-dimensional memory array which is field programmable is described. In the following description, numerous specific details are set forth such as specific materials and layer thicknesses. It will be apparent, however, to one skilled in the art that the present invention may be practiced without these details. In other instances, well-known circuits and fabrication techniques have not been set forth in detail in order not to unnecessarily obscure the present invention.

OVERVIEW OF THE STRUCTURE OF THE INVENTED MEMORY ARRAY

The invented memory array is fabricated on several levels and, for instance, may have eight levels of storage. Each level includes a first plurality of parallel spaced-apart rail-stacks running in a first direction and a second plurality of rail-stacks or conductors (depending on the embodiment) running in a second direction. Generally, the first rail-stacks run perpendicular to the second conductors/rail-stacks and hence form a right angle at their intersections. (In the invented array as well as in the prior art, conductors at one level are shared with the next level, hence the term "level" may not be precisely descriptive.)

The use of rail-stacks is a departure from prior art three-dimensional memories where conductors alone were used in lieu of rail-stacks, and

where discrete cells (e.g., pillars) were formed at the intersections of the lines. As will be seen, a bit is stored at each of the intersections of rail-stacks. However, there is no apparent individual memory cell at the intersections, rather memory cells are defined by the rail-stacks and intermediate layers. This makes it easier to fabricate the invented array as will be seen. When the array is fabricated all the bits are in the zero (or one) state and after programming, the programmed bits are in the one (or zero) state.

In the embodiment Figure 1 several rail-stacks are illustrated in the partial cross-section of the invented array. For instance, rail-stack 16 is shown at one height and a half rail-stack 18 is shown at a second height above the first height. Also, half rail-stacks are disposed between rail-stack 16 and a substrate 10. These lower rail-stacks run in the same direction as the half rail-stack 18. A bit is stored at the intersection of rail-stacks and, for instance, a "cell" is present between the rail-stacks and layers shown within the bracket 17 and another within the bracket 19. Each of these brackets span a memory level.

The array is fabricated on a substrate 10 which may be an ordinary monocrystaline silicon substrate. Decoding circuitry, sensing circuits, and programming circuits are fabricated in one embodiment within the substrate

10 under the memory array using, for instance, ordinary MOS fabrication techniques. (These circuits may also be fabricated above the substrate.) Vias are used to connect conductors within the rail-stacks to the substrates to allow access to each rail-stack in order to program data into the array and to read data from the array. For instance, the circuitry within the substrate 10 may select rail-stack 16 and the rail stack 18 in order to either program or read a bit associated with the intersection of these rail-stacks. (In the case of the embodiments of Figure 5 some conductors are not part of rail-stacks; these conductors are also coupled to the substrate circuits.)

As shown in Figure 1, an insulating layer 12 is formed over the substrate in order that the array may be fabricated above the substrate. This layer may be planarized with, for instance, chemical-mechanical polishing (CMP) to provide a flat surface upon which the array may be fabricated.

Following this, a conductive layer 14 is formed on the substrate. As will be seen, conductive layers are used within the rail-stacks and these layers and the resultant conductors may be fabricated from elemental metals such as tungsten, tantalum, aluminum, copper or metal alloys may be used such as MoW. Metal silicides may also be used such as $TiSi_2$, $CoSi_2$ or a conductive compound such as TiN, WC may be used. A highly doped

semiconductor layer such as silicon is also suitable. Multiple layer structures may be used selecting one or more of the above.

Following the deposition of a conductive layer, a layer of semiconductor material (layer 15) such as silicon is formed over the conductive layer. This is typically a polysilicon layer, however, an amorphous layer may be used. Other semiconductor materials may be used such as Ge, GaAs, etc. In the embodiment of Figure 1 this semiconductor layer is highly doped and, as will be seen, forms one-half a diode. After masking and etching steps, half rail-stacks are formed. These rail-stacks are "half" or partial rail-stacks since they are approximately half the thickness of the rail-stacks used in the next level.

Following this, in the embodiment of Figure 1, a material for the antifuses used to program the array is deposited. In one embodiment, the layer 20 is a dielectric such as silicon dioxide which is deposited by chemical vapor deposition (CVD) in a blanket deposition over the half rail-stacks and a dielectric fill, filling the space between the rail-stacks. In another embodiment the layer 20 is grown on the upper surface of the silicon layer 15 and only exists on the rail-stacks.

Now a full set of memory array rail-stacks is formed on the layer 20. This comprises first the deposition of a lightly doped silicon layer 21 doped

with a conductivity type dopant opposite to that used for the silicon layer 15, a heavily doped silicon layer 22 doped also opposite to the layer 15, a conductive layer 23 and a heavily doped silicon layer 24 doped with the same conductivity type dopant as layers 21 and 22. After masking and etching, the rail-stacks shown in Figure 1, such as rail-stack 16 are formed. These rail-stacks are, as illustrated, in a direction perpendicular to the rail-stacks above and below them.

While not shown in Figure 1 but as will be described later, the spaces between the rail-stacks after they are defined, are filled with a dielectric such as silicon dioxide. Then the rail-stacks and fill are planarized by CMP. In another embodiment spin-on-glass (SOG) is used to fill the voids, in this case chemical planarization can be used, for example, plasma etching. Other fill and planarization methods can be used.

After formation of the rail-stacks another antifuse layer 26 is formed, for instance, from a dielectric such as silicon dioxide, silicon nitride, silicon oxynitride, amorphous carbon or other insulating materials or combinations of materials. (Also an updoped layer of silicon may be used for the antifuse layer.)

Now another layer of rail-stacks are defined and only half rail-stacks are shown in Figure 1 at this upper level. This half rail-stack comprises a

silicon layer 28 doped with a conductivity type dopant opposite to that of layer 24. This is a lightly doped layer. Another silicon layer 30 is formed on layer 28 and this layer is doped with the same conductivity type dopant as layer 28, however, it is more heavily doped. Then a conductive layer 31 is formed above the layer 30.

Half rail-stacks are used at the very upper-most level of the array and at the very lowest level of the array. In between the half rail-stacks the full rail-stacks, such as rail-stack 16, are used throughout the array.

It should be noted that the silicon layers disposed on the conductive layers extend the entire length of the rail-stacks in the embodiment of Figure 1 and are uninterrupted except possibly where vias are used to provide a conductive path to the substrate 10.

In Figure 1 a path 32 is illustrated from a lower conductor in level 17 to an upper conductor in this level found in the rail-stack 18. This path is accessed in one embodiment through decoding circuitry in the substrate for both programming and reading of data into and from the array for one bit.

For instance, to program the bit, a relatively high voltage, 5-20V is applied between the conductors generally so as to forward-bias the "diode" between these conductors. This relatively high voltage causes a breach in the layer 26 creating a diode. Without this high voltage, the layer 26 remains

an insulator. Thus, by selecting pairs of conductors, diodes can be selectively formed so as to program the array. While programming the array with the layers adjacent to the antifuse material being forward-biased is currently preferred, it is also possible to program using a reverse-biasing potential.

To sense the data programmed into the array, a voltage lower than that for programming is used. This voltage is applied so as to forward-bias the diode of the cell being accessed and thus allowing a sense amplifier to determine whether or not the layer 26 is intact between the rail-stacks. Note that "sneak" or parasitic paths in the array which would interfere with the sensing will include a reverse-biased diode.

Also as will be described later, the "anode" and "cathode" of the diodes are reversed at each of the successive antifuse layers. This facilitates easier programming and sensing since all of its conductors at each level are either bit lines or word lines. And, for instance, conductors at one height will serve as bit lines for two levels and conductors at the next height serve as word lines for two levels. This simplifies the decoding and sensing and more importantly reduces processing.

EMBODIMENT OF FIGURE 3

In the cross-section elevation view of Figure 3, one embodiment is illustrated which corresponds to the embodiment shown in Figure 1. In Figure 3 the half rail-stacks of Figure 1 are not illustrated. Three complete levels 35, 36 and 37 of the array are illustrated in Figure 3. Below layer 38 of Figure 3 other rail-stacks or half rail-stack are used. Also above layer 65, a full or half rail-stack is used.

The rail-stack 3 comprising layers 38 through 41 includes a lightly doped n- layer 38, a heavily doped n+ layer 39, a conductor layer 40 and n+ layer 41. The fabrication of these rail-stacks will be discussed in more detail in conjunction with Figure 2A through Figure 2G. An antifuse layer 42 which for the embodiment of Figure 3 is a blanket deposition covers all of the rail-stacks formed below layer 42 as well as the fill filling the voids between the rails. As mentioned, the layer 42 is a deposited silicon dioxide layer in one embodiment.

It should be noted that n+ layers sandwich the conductor layer 40. These highly doped layers provide ohmic transitions to prevent unintended Schotky formation.

The layers above and below conductor 40 are not symmetrical for the embodiment illustrated in that an n- layer 38 is used below the conductor 40 and not above the conductor 40. Only a single lightly doped layer (in

conjunction with a heavily doped layer) is needed to define a diode; the thickness of this lightly doped layer is important in controlling the break-down voltage and resistance of the diode so formed. The layer 41, a heavily doped semiconductor layer, and the fill are planarized after the rail-stacks are defined and then a blanket deposition of the antifuse layer 42 is formed on the layer 41. (The lines 43 in Figure 3 are used to indicate that the antifuse layer 42 and like layers are not etched with the rail-stack below it and thus extend over the entire array for the illustrated embodiment.)

One advantage to the layer 42 and the other like layers in the structure, such as layers 51, 56 and 65, is that since they are an unbroken deposition, sidewall leakage (into the rail-stacks below) will be minimized, limiting electrical problems during reading and writing. When subsequent conductive material is deposited, it is unable to reach the sides of the rail-stacks below it because of this blanket deposition of the antifuse layer. For instance, path 49 which would allow silicon from layer 52 to cause a parasitic path does not exist because of the unbroken blanket deposition of the antifuse layer 51.

Rail-stacks 4 comprising layers 44, 45, 46 and 47 are formed on the antifuse layer 42. Layer 44 is lightly doped with a p-type dopant for the embodiment illustrated followed by a p+ layer 45, a conductive layer 46 and

a p+ layer 47. After these layers are deposited, they are masked and etched to define the rail-stacks. Then the voids between these rail-stacks, such as void 50, are filled with a dielectric. The fill dielectric is planarized along with a portion of p+ layer 47. Planarization is done at this point in the fabrication since there is generally poor control over the thickness and contour of the fill. The fill tends to build up on the rail-stacks when a non-spin-on deposition is used. This is followed by a blanket deposition of layer 54.

The process is now repeated this time beginning with an n- layer 52 followed by an n+ layer 53, a conductive layer 54 and n+ layer 55. Again after defining the rail-stacks 5, the voids are filled and the surface is planarized. Another antifuse layer 56 is deposited.

The process is repeated for the rail-stacks 6 this time beginning with a p- layer 61, p+ layer 62, conductive layer 63, p+ layer 64. Again after defining the rail-stacks, filling the void 60 and then planarizing, another antifuse layer 65 is deposited.

As shown by the path 66, when a large enough voltage is applied between conductors 46 and 54 the antifuse layer 51, at the intersection of layers 47 and 52 is breached creating a diode at the intersection. As mentioned, this is selectively done throughout the array to program the array. The conductor 54 is therefore a bit line for the "cells" above and

below it, for instance path 67 indicates another possible current path for another "cell" where the conductor 54 is again a bit line during sensing.

It should be noted that with the reversal of the p- and n- layers at each successive rail-stack, planarization always occurs on a heavily doped layer such as layer 47 and layer 55. Moreover, the lightly doped layers are always formed on relatively planar surfaces, consequently their thickness can be more easily controlled. This, as mentioned, allows the characteristics of the diode (once the intermediate antifuse layer is breached) to be more reliably controlled.

PROCESSING FLOW FOR THE EMBODIMENT OF FIGURE 3

The process flow for forming rail-stack 5 of Figure 3 is illustrated in Figures 2A-2H. It will be apparent that the rail-stacks for the other embodiment (Figures 4 and 5) are similarly processed.

First, as shown in Figure 2A an antifuse layer 51 is deposited. This typically is 50-200Å of silicon dioxide which can be deposited with any one of very well-known processes. Following this, a silicon layer 52 is deposited which is typically 1000-4000Å thick and formed with a CVD process where a phosphorous dopant is deposited along with the deposition of for instance, the polysilicon semiconductor material or where the dopant is ion implanted

following the deposition of the layer. This layer is doped to a level of 5×10^{16} - 10^{18}/cm^3.

Now, as shown in Figure 2B an n+ layer 53 is deposited again using CVD. This layer may be approximately 300-3000Å thick and in one embodiment is doped to a level of $>10^{19}$/cm^3.

Throughout this application two adjacent silicon layers are often shown such as layers 52 and 53, with different doping. These layers may be formed with one deposition and then using ion implantation step at two different energy levels to obtain the two doping levels.

A conductive layer which may be 500-1500Å thick is formed using any one of numerous well-known thin film deposition process such as sputtering. A refractory metal may be used or a silicide of a refractory metal. Also as mentioned aluminum or copper can be used, or more simply the heavily doped silicon can be the conductor.

Next another semiconductor layer of, for instance, polysilicon approximately 1500-2000Å thick is formed again doped to a level of $>10^{19}$/cm^3. This is shown as layer 55 in Figure 2D; after planarization its thickness is between 300Å and 2000Å thick.

A masking and etching step is now used to define rail-stacks, such as rail-stacks 69, 70 and 71 shown in Figure 2E. Note that when comparing

this view to the view of rail-stack 5 of Figure 3, the view in Figure 2E is taken from the side and consequently shows the individual rail-stacks. An ordinary masking and etching step for instance using plasma etching, may be used. Etchants can be used that stop on the antifuse layer thus preventing this layer from being etched away. Thus, layer 51 can be considered an etchant stop layer depending on the specific etchants used.

Now as shown in Figure 2F, the spaces between the rail-stacks are filled with a dielectric such as formed with a HDPCVD process.

Chemical-mechanical polishing is then employed to planarize the upper surface of the rail-stacks shown in Figure 2F in one embodiment. Chemical etching can also be used as mentioned with certain dielectrics. This planarization can reduce the thickness of the layer 55 to approximately 500Å, thus this layer ends up being of approximately the same thickness as the layer 53.

Next as shown in Figure 2H another antifuse layer 56 is formed on the planarized surface 75. Since the layer 56 is deposited over all the rail-stacks and the filler material and remains unetched, it forms a barrier to the migration of the materials subsequently deposited that might make their way along the sides of the rail-stacks such as along path 79. Thus the layer 56

helps prevent the parasitic paths and potential shorts that may occur with prior art memories.

It should be noted that in Figure 3 while the antifuse layer is shown as a blanket layer covering the rail-stacks and fill, it is possible also to fabricate each level where the antifuse layer is in fact grown from a semiconductor layer. For instance, an oxidation step may be used to grow a silicon dioxide layer from layers 41, 47, 55 and 64. This grown layer would then be in lieu of the antifuse layers shown in Figure 3.

THE EMBODIMENT OF FIGURE 4

For the embodiment of Figure 4 each rail-stack begins with a conductor such as layer 80 of Figure 4. An n+ semiconductor layer 81 and an n- layer 82 are formed on layer 80. Next a layer of antifuse material 83 is formed. Then a p+ layer 84 of semiconductor material is deposited (e.g., silicon with boron dopant) on the antifuse. When the rail-stacks are formed, for instance for rail-stack 2 of Figure 4, the antifuse layer 83 is etched as well as layers 80, 81, 82 and 84.

The voids between the rail stacks are now filled and planarization is done, planarizing the fill with the upper surface of the layer 84. Following the completion of the rail-stack 2 the next rail-stacks are formed shown as rail-stacks 3 in Figure 4. This comprises a conductor layer 85, p+ layer 86, p-

layer 87, antifuse layer 88 and n+ layer 89. Again masking and etching occur. This etching also etches the exposed regions of layer 84 which does not appear in the view of Figure 4, but this will be apparent shortly when region 95 of the next stack is discussed. Now filling and planarization occurs and the next layer of rail-stacks are formed shown as rail-stack 4. As illustrated, this comprises a conductive layer 90, n+ layer 91, n- layer 92, antifuse layer 93, and p+ layer 94. Once again masking, etching, filling and planarization occur.

Unlike the embodiment of Figure 3, when rail-stacks at any particular height are formed, etching must occur on one layer of the rail-stack immediately below the rail-stack being defined. For instance, when rail-stack 4 is etched the layer 89 of rail-stack 3 is etched away where it is not covered by rail-stack 4 as shown by region 95. This etching is used to remove all of the semiconductor material between the adjacent conductors and consequently prevent a path, such as path 96 shown in Figure 4. This etching also occurs to layer 84 which, as mentioned, is not seen in Figure 4. In this connection the antifuse layer 88 can be used as an etchant stop, although this is not necessary. No harm is done if etching does occur through the layer 88 since the antifuse layer is only needed at the intersections of the rail-stacks. Note the etching of the region 95 is done in

alignment with overlying rail-stacks and consequently no additional masking is required.

As was the case with the earlier embodiment, the order of the n and p doped layers alternate with each successive rail-stack. Moreover, the rail-stacks at any given height include both p and n layers. In contrast, for the embodiment of Figure 3, at any particular height, the rail-stacks are doped with either an n type or p type dopant but not both.

EMBODIMENT OF FIGURE 5

In the embodiment of Figure 5, alternate levels of rail-stacks running in a first direction and intermediate layers of conductors are running in a second direction are used. For instance as shown in Figure 5, the conductors 3, 5 and 7 run in a first direction whereas the rail-stacks 4 and 6 run in a second direction.

In this embodiment each of the rail-stacks is symmetrical about a conductor such as conductor 109 of rail-stack 4. The conductor is sandwiched between two n+ layers 108 and 110. More lightly doped outer layers 107 and 111 are disposed on these more heavily doped layers.

In fabrication the conductors such as conductors 105, are first formed, for instance, on the substrate. The spaces between these conductors may be filled and planarization may occur. Then an antifuse layer 106, n- layer

107, n+ layer 108, conductive layer 109, n+ layer 110 and n- layer 111 are deposited. Rail-stacks are then defined by masking and etching. The voids between the rail-stacks are then filled with a dielectric. Then planarization of the filling material and the upper surface of layer 111 is performed. Following this, antifuse layer 112 is deposited over the entire array. Now additional conductors are formed such as conductors 113. Each level in this array is between a metallic conductor such as conductor 105, and a sandwich conductor such as conductor 109. Thus there are four memory levels shown in Figure 5, levels 100, 101, 102 and 103.

Programming in this array causes the formation of Schottky diodes consequently, the conductors such as conductors 105 and 113 must be of a suitable material to allow formation of a Schottky diode. For instance, aluminum and some refractory metal or silicides may be used.

OTHER EMBODIMENTS

In the above description a conductor is shared by two levels. An array may be fabricated where there are two conductors for each level that are not shared with other levels. A dielectric may be used to separate each such level. Also while above diodes on alternate levels "point" in the same direction, this is not necessary. For instance, a shared conductor may have

diodes point-in from above and point-out from below. This requires different driving circuitry in the substrate.

All the above embodiment have benefits over the prior art three-dimensional memories. One advantage is that the diodes are formed by breaching an antifuse layer. This results in diodes with very small junction areas. The resultant low-leakage diodes improves the performance of the array. Additionally, etching is not as deep as with the prior art three-dimensional memories. Difficulties with stringers where individual pillars were used in the prior art is eliminated with some of the above embodiments. The three embodiments provide numerous material choices and "post-write diode" choices.

Thus a three-dimensional memory array has been described using rail-stacks which simplifies processing and provides better performance over prior art three-dimensional arrays.